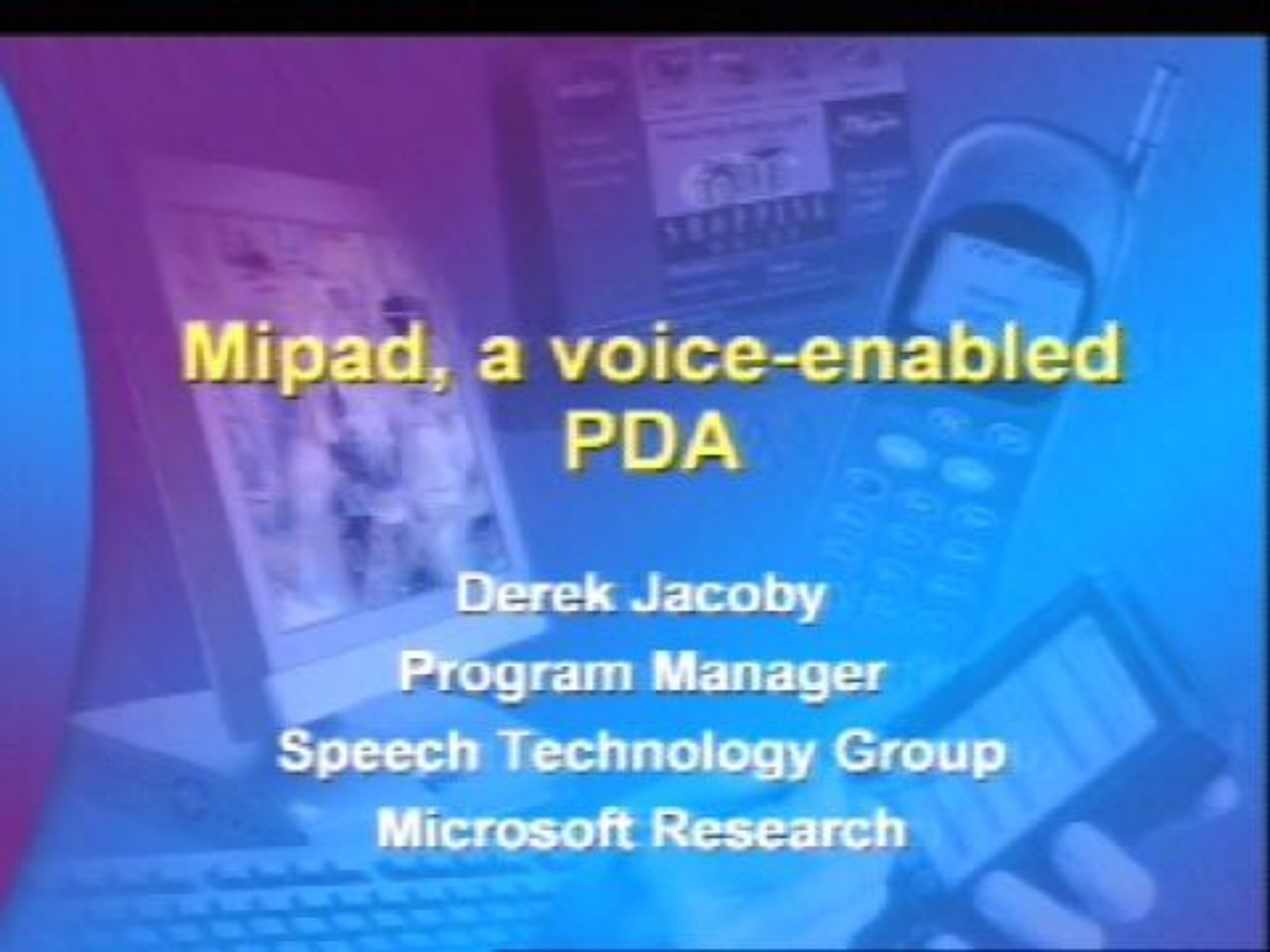

Mipad, a voice-enabled PDA
Derek Jacoby
Program Manager
Speech Technology Group
Microsoft Research

Agenda

- What is the Speech Technology Group (STG) ?
- Introduction to Mipad
- Mipad Demo!
- Technologies in Mipad
- Mipad future plans
- Speech in your products

STG Mission

- *Develop best-of-breed spoken language engine technologies*
- *Develop compelling new interaction models*
- *Help to integrate speech into MS applications and Windows*

STG Overview

- People - 17 full time, 5 intern/contract
  - 10 researchers and senior researchers
  - 5 developers
  - 1 test lead
  - 1 program manager

STG History


•Started
•Whisper CSR, SAPI 1.0
•SAPI 2.0, Phone 1.0
•Whistler TTS, SAPI 3.0, Agent
•SPAI 4.0, NT 5.0, Phone 2.0, Encarta
•Transferred 15 FTEs to IIT
•Beefed up research
Launched Dr. Who
•Launched Mipad

Overview of Projects

- Mipad
  - Designed for Pocket PC devices
  - Email and scheduling tasks
  - Currently relies on high bandwidth network connection
- Unified engine
- Noise robustness
- Dialog management
- Recognition on smaller devices

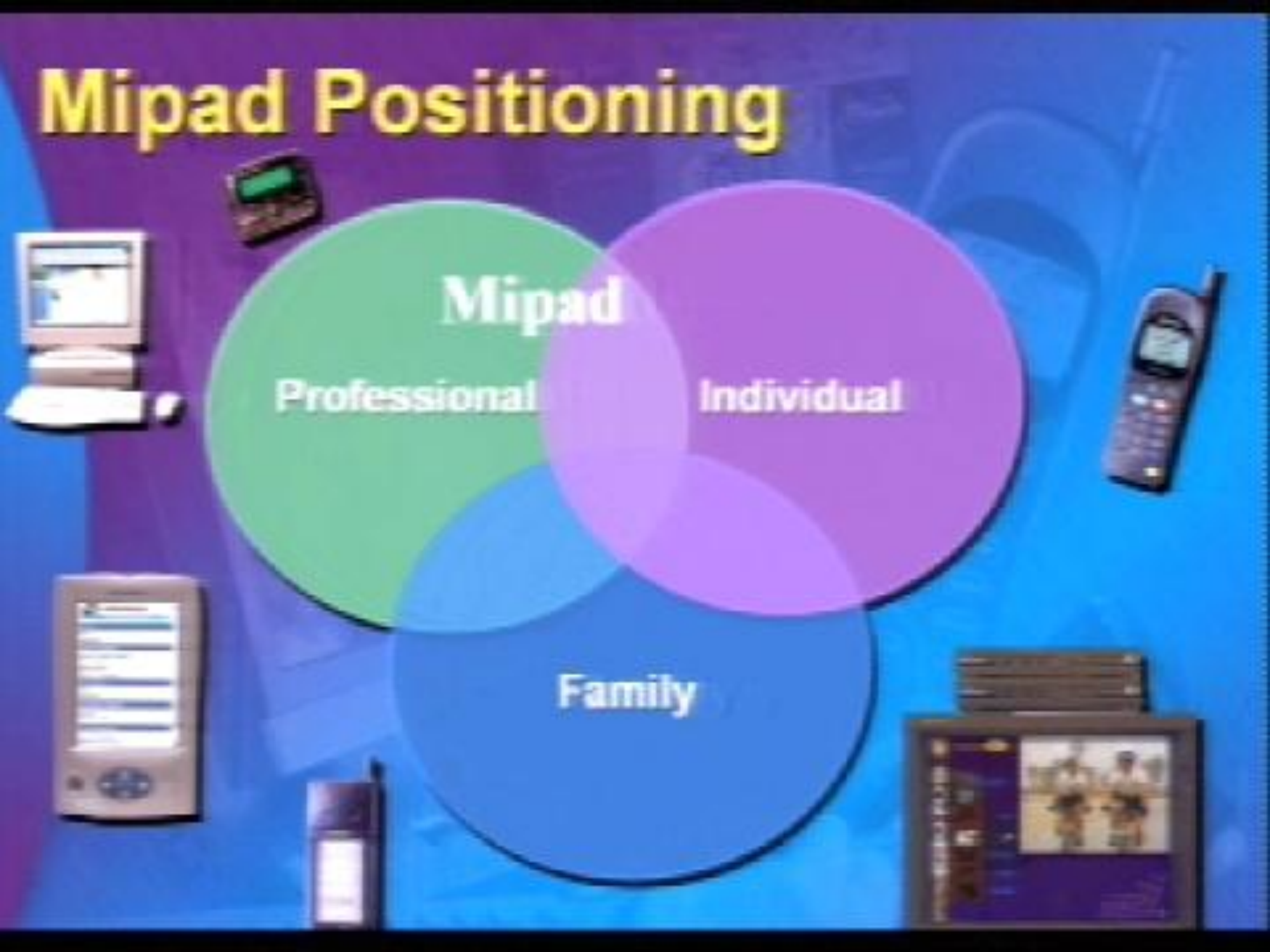
Mipad

- Microsoft Interactive notePAD
- Demonstration application
  - More than a prototype, less than a product
  - Planned for extensive internal use at MS
  - Software project, but will be available (internally) as a bundled app with Pocket PC

Mipad Goals

- Create a compelling spoken interaction model for PDA tasks
- Identify speech research required to make spoken user interfaces compelling
- Collect real-world data on speech usage
- Transfer the technology and ideas to the product groups

Mipad Positioning



MiPad
Our 1st Dr Who Application

- *Problems need to be fixed:*
  - *Entering text is very slow*
  - *Small screen means functions are hidden*
- *Good ergonomics for speech+pen*
  - *Two complementary modalities*
- *Multimodal interface enhances usability*
  - *Tap-driven CSR and SLU*
  - *Overcome today's SLU technology*

Mipad Features

- Email tasks
  - Send/read new email
  - View inbox and other folders
  - Limited search and filtering
- Schedule Tasks
  - New Appointment
  - View calendar
- Address book/contact list
  - Look up a person by name
- Voice browser tasks
  - MSR Beijing collaboration (Chang)

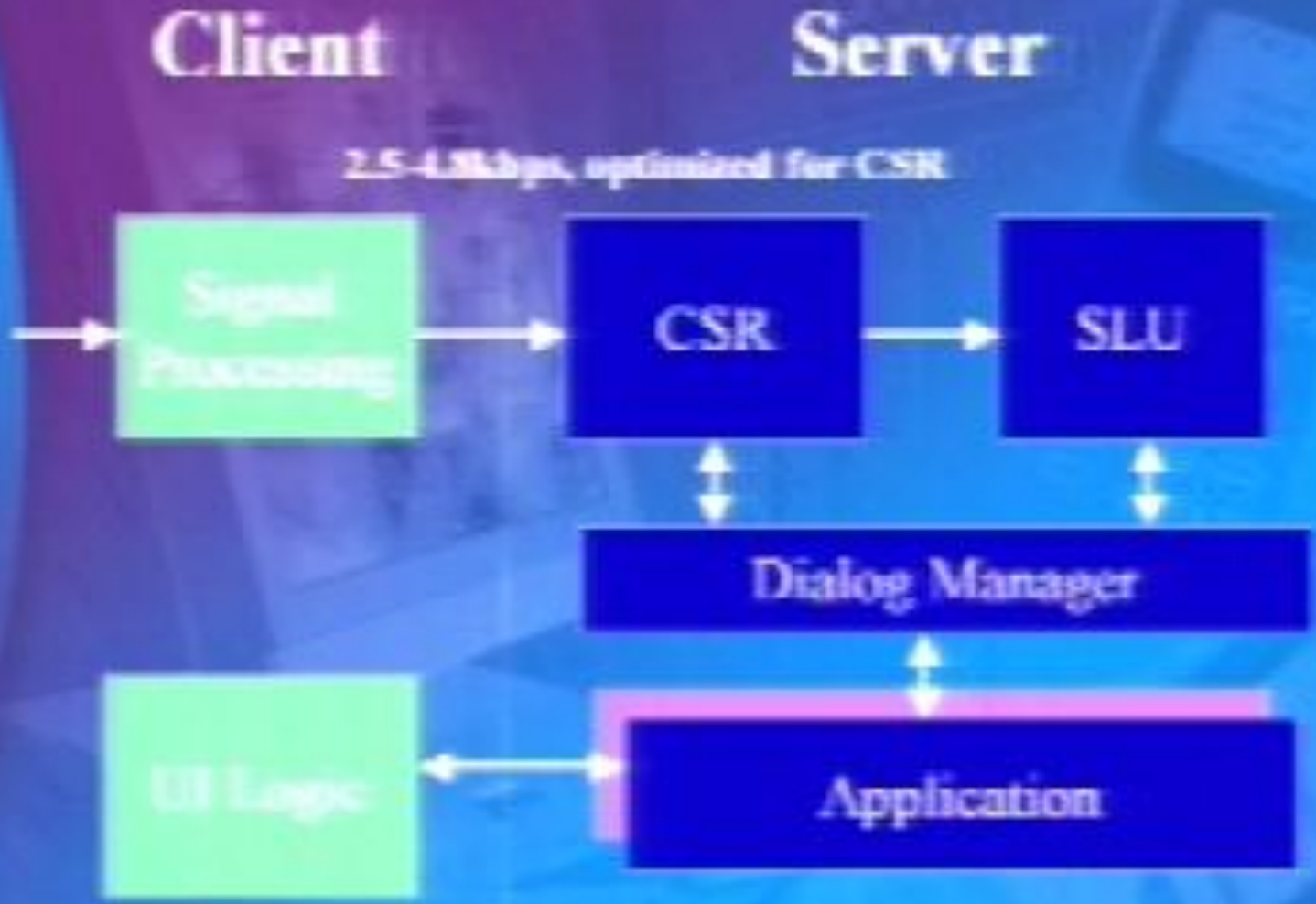
Mipad Schedule

- STG internal trials start in September
- Release to Microsoft Research planned for December
- Availability on MS internal order system for bundling with new Pocket PC devices by Q1 CY'01

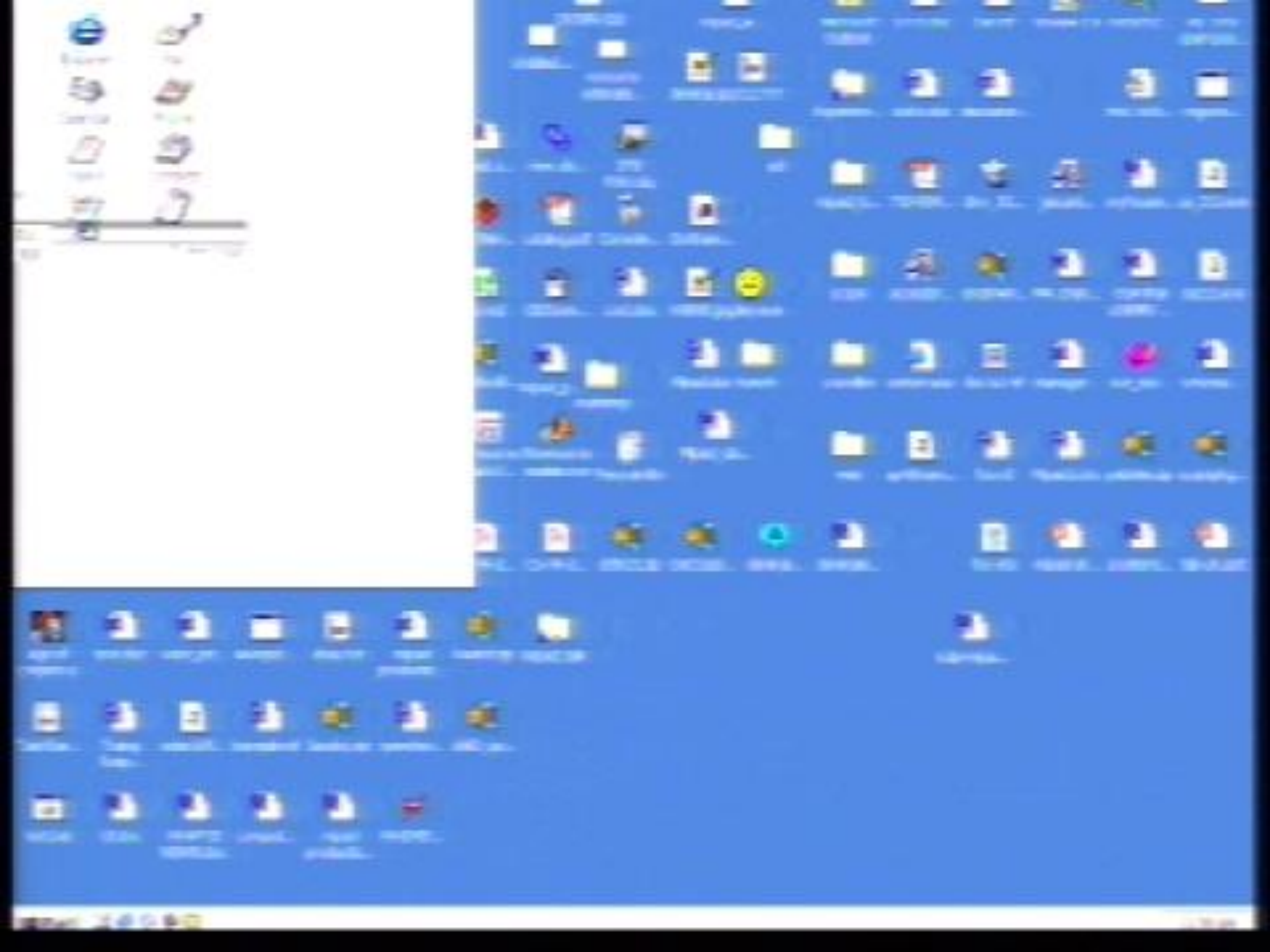
Mipad Technologies

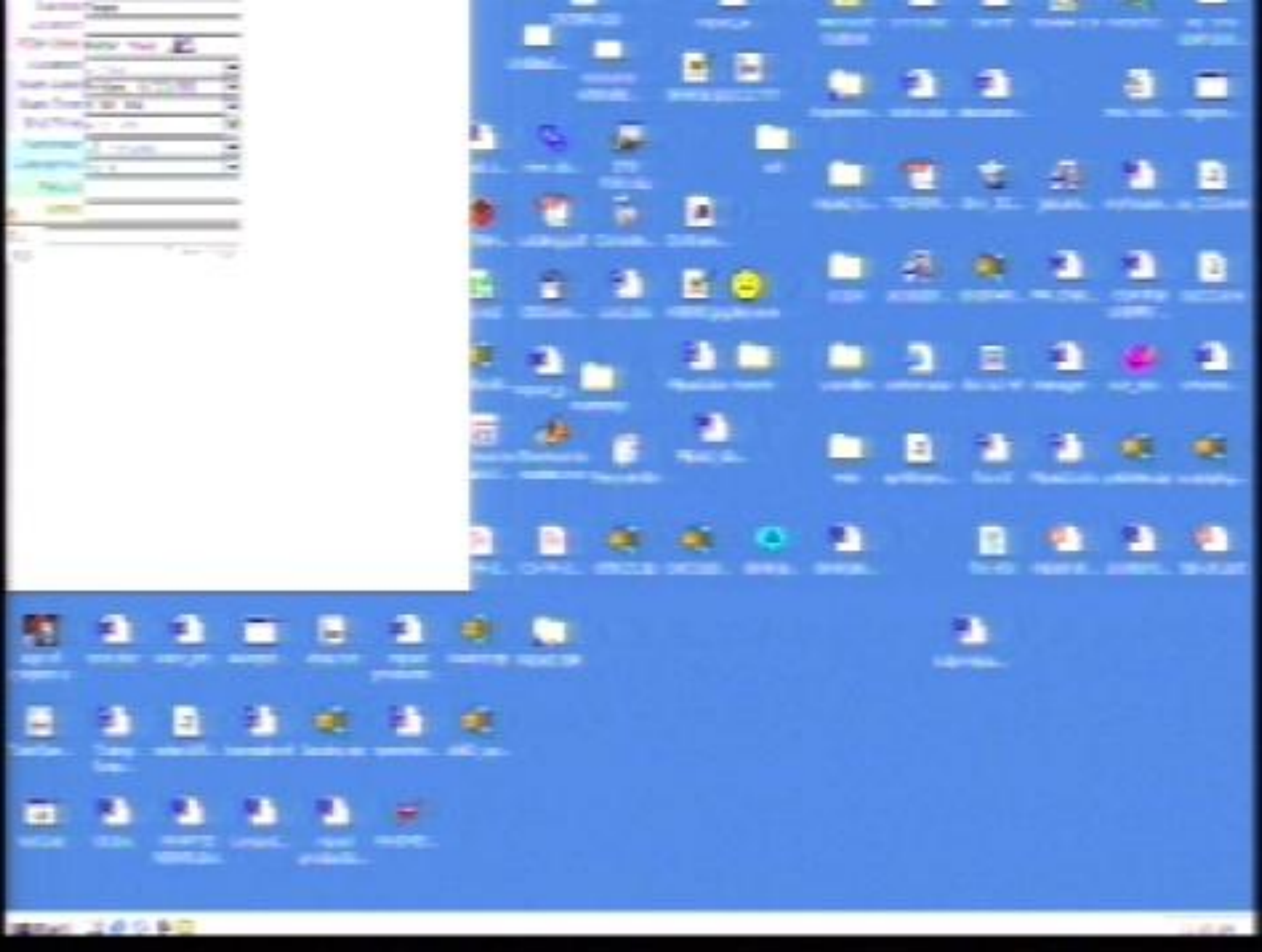
- As a research demonstration application, the technologies Mipad uses are the focus
  - Unified Engine
  - Noise Robustness Algorithm
  - Dialog Management
- The user interface is also a focus
  - Flexible interaction model
  - Methods for evaluating speech user interfaces
  - Guidelines for product groups

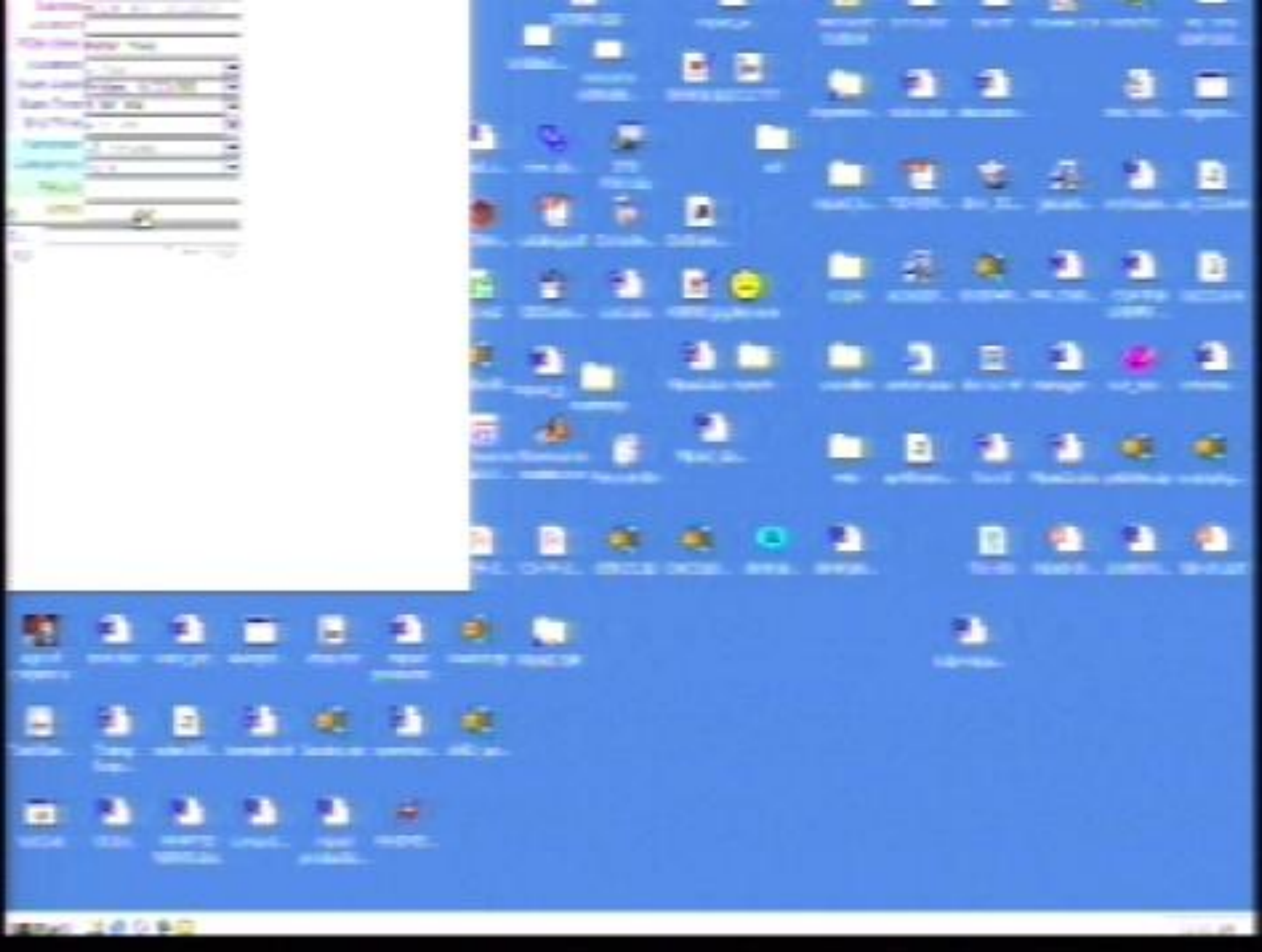
MiPad's Architecture

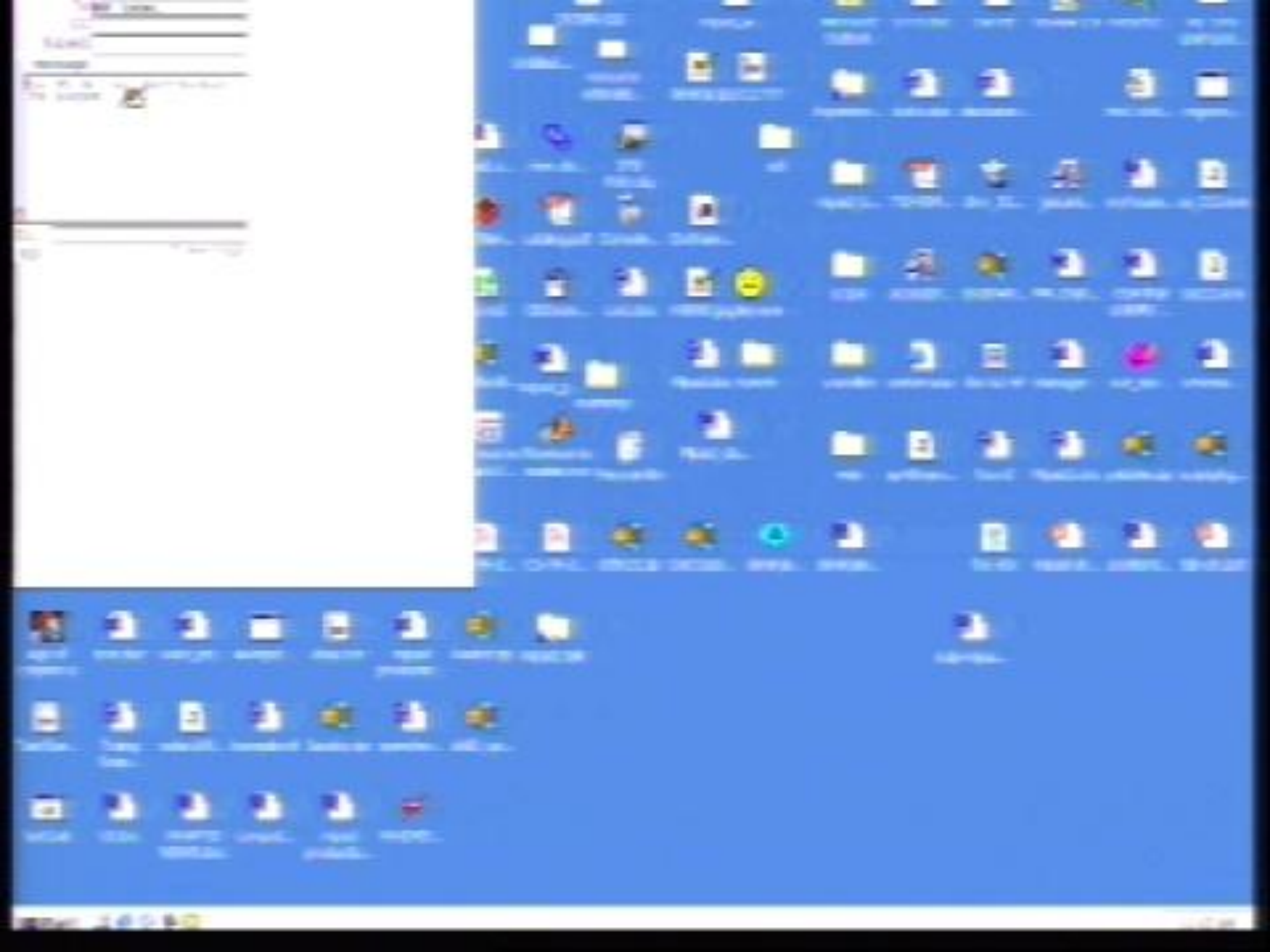


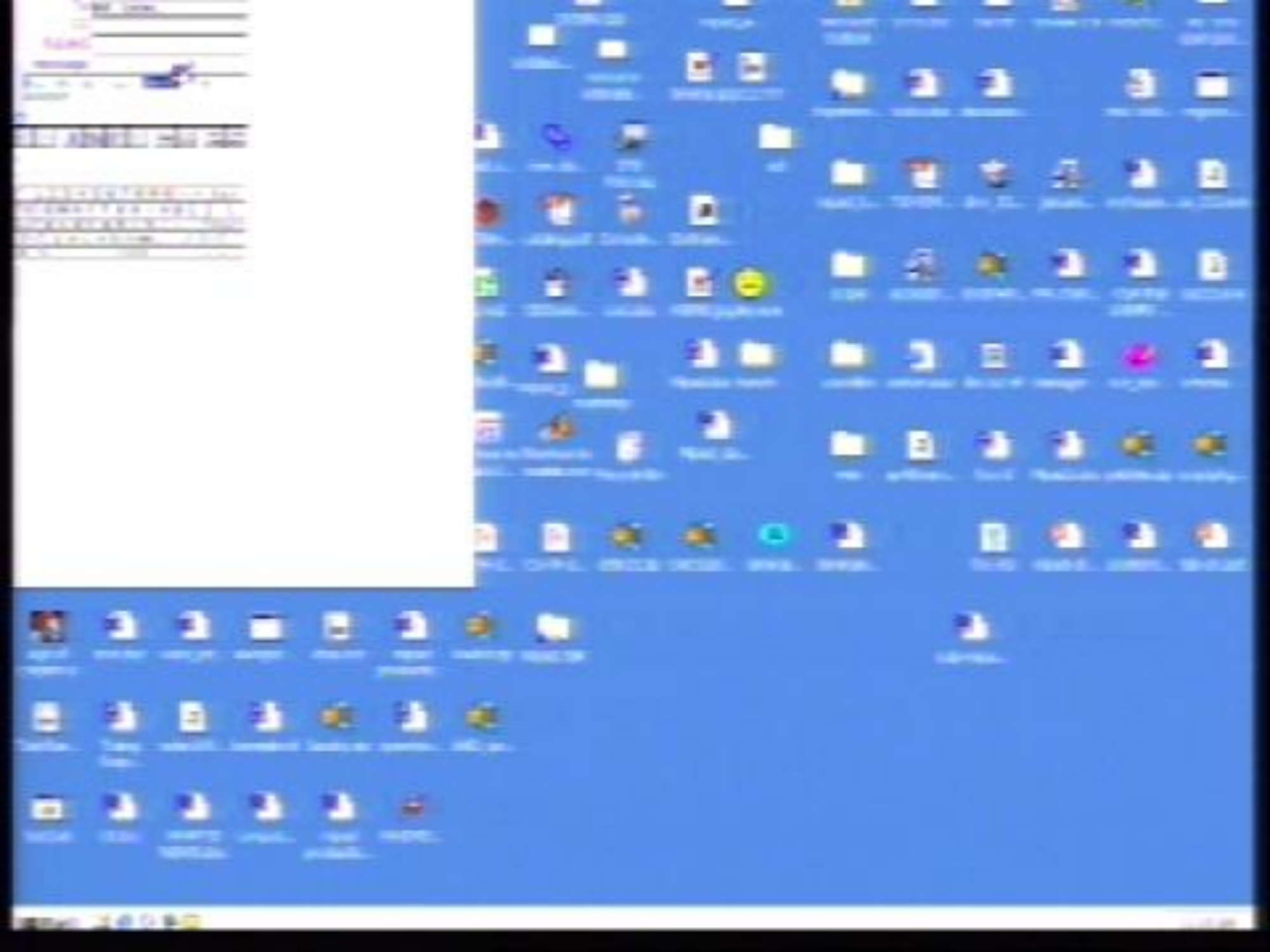
Mipad Demonstration!

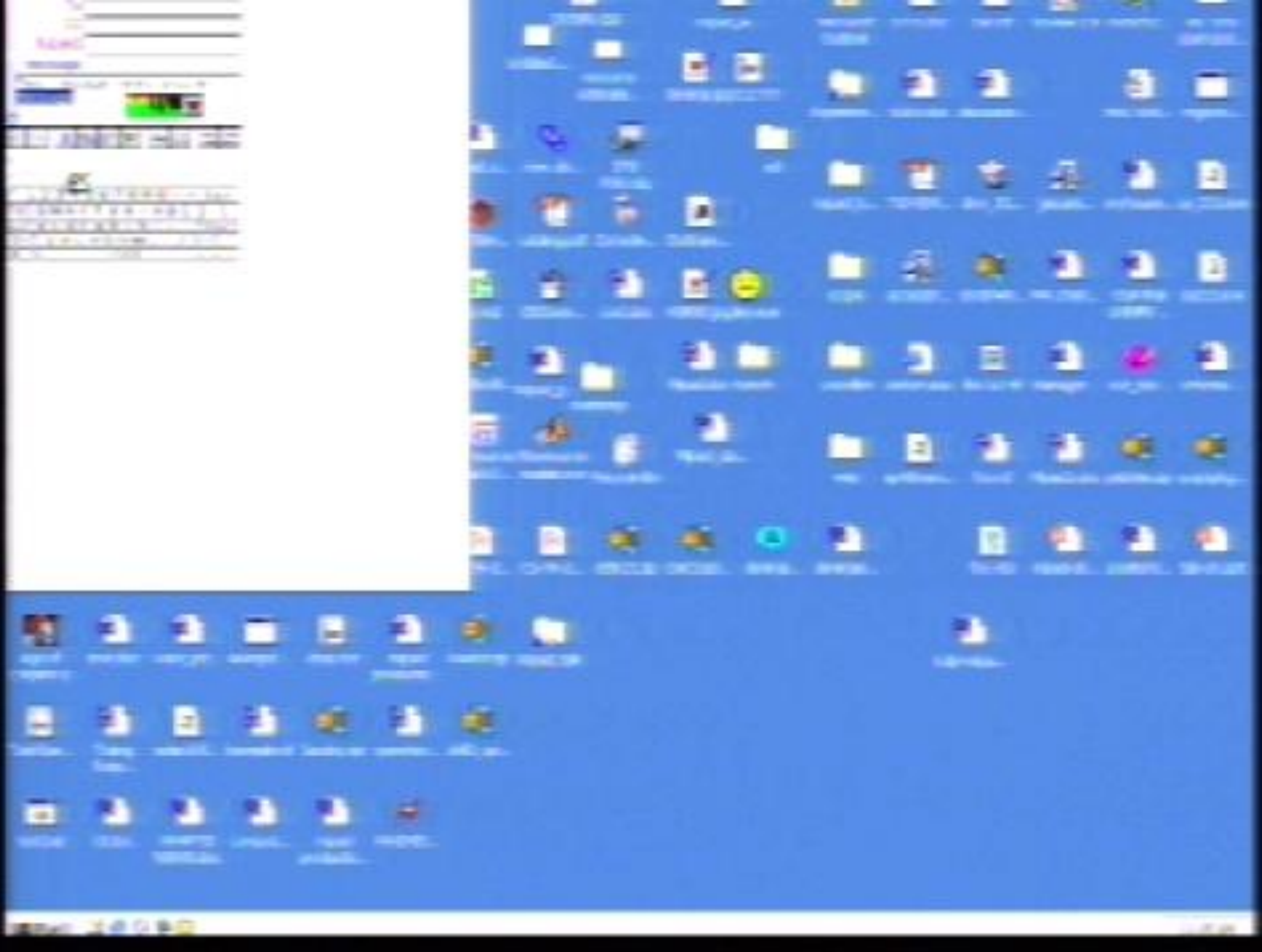


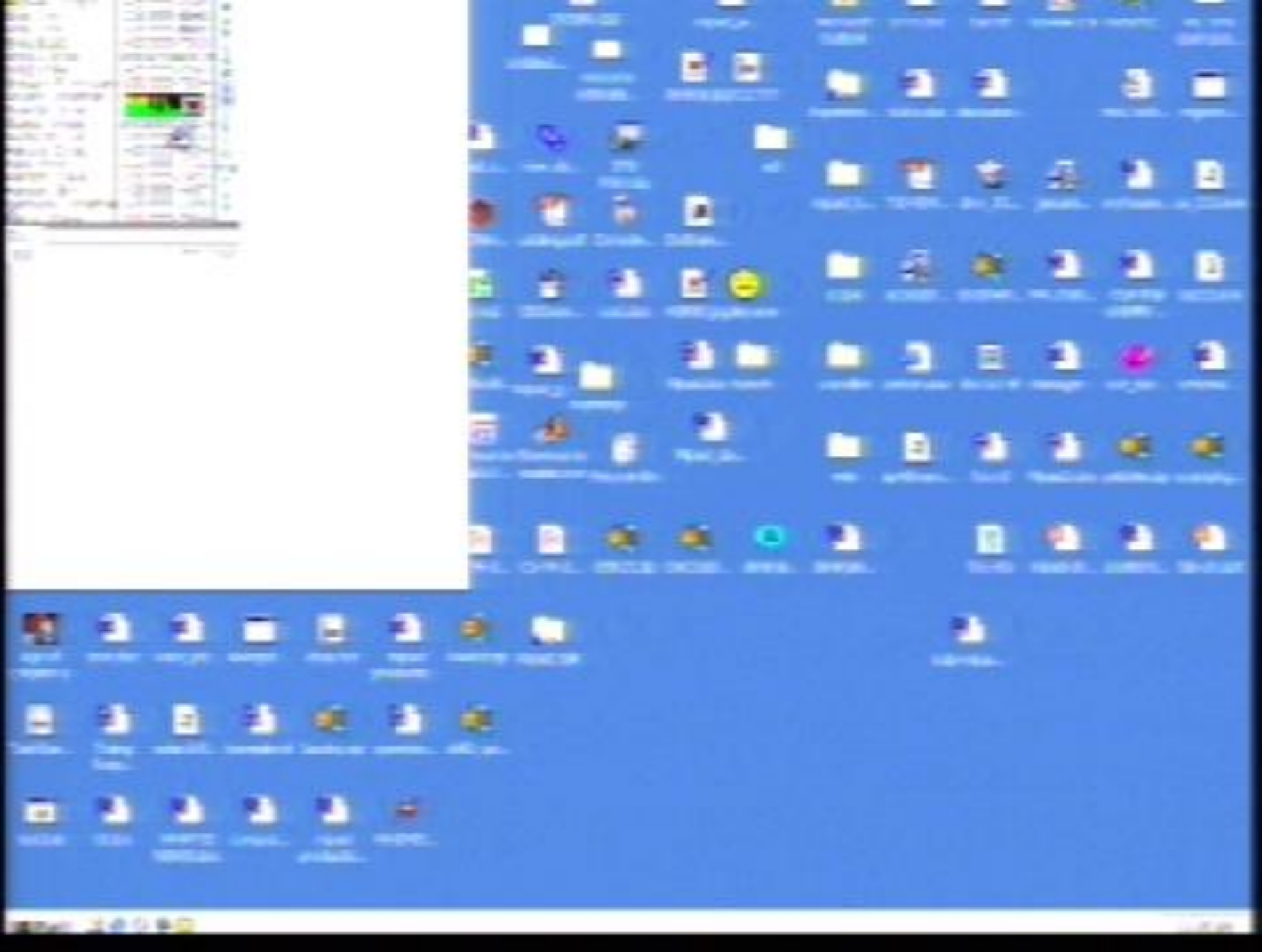












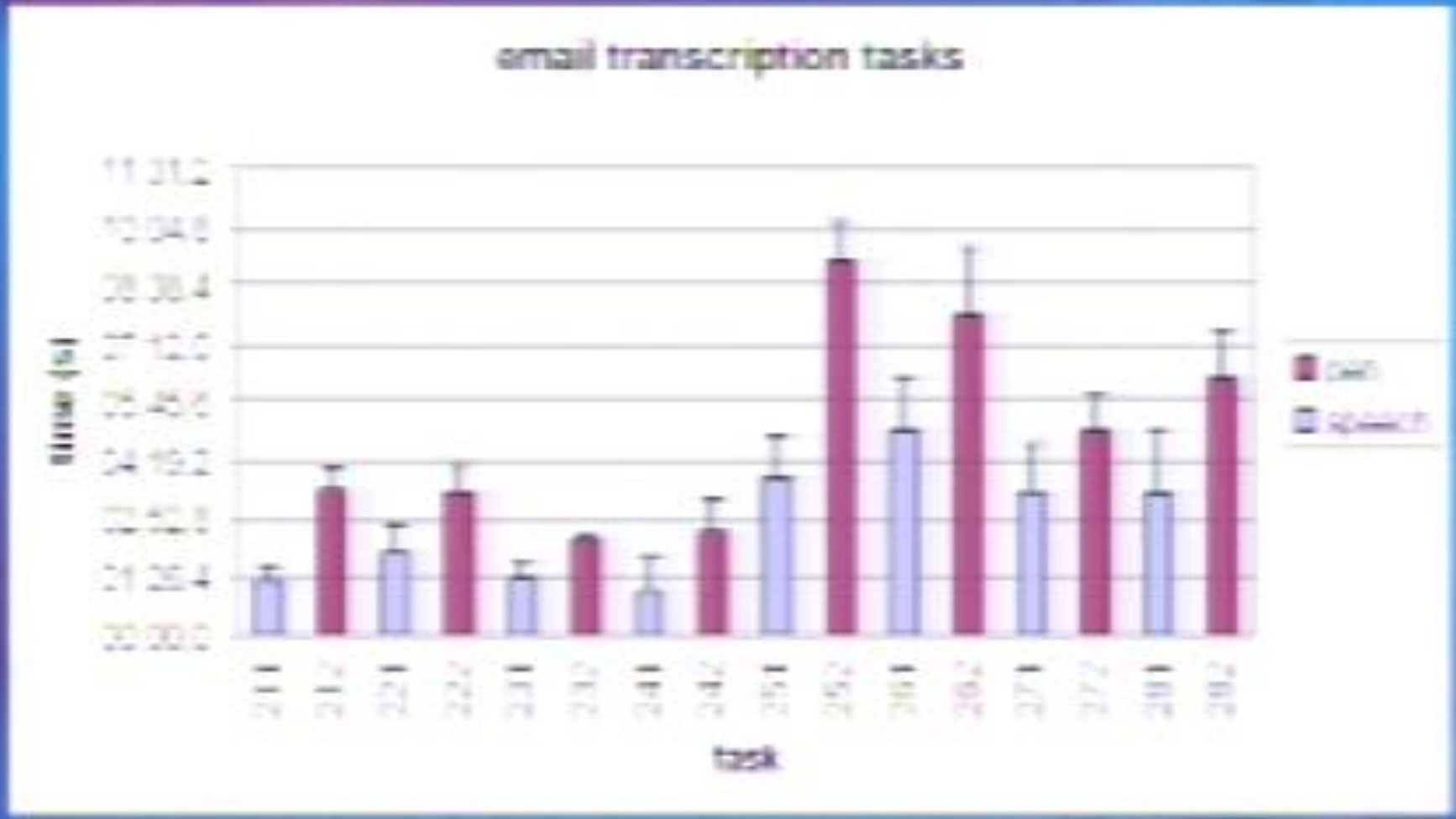
Mipad Interface

- Flexible interaction model
- Methods for evaluating speech user interfaces
- Guidelines for product groups
- Longitudinal user study
  - Track real users for a long period of time
  - Collect real data on real devices
  - Using instrumented MiPad

Preliminary Benchmark

- 48 tasks
  - Each subject gets 24 tasks
  - ½ using pen and ½ using speech
- MiPad Throughput
  - Overall 30% faster than PPC
  - Email transcription task is 2 times faster
- Subjective ratings
  - User experience very positive

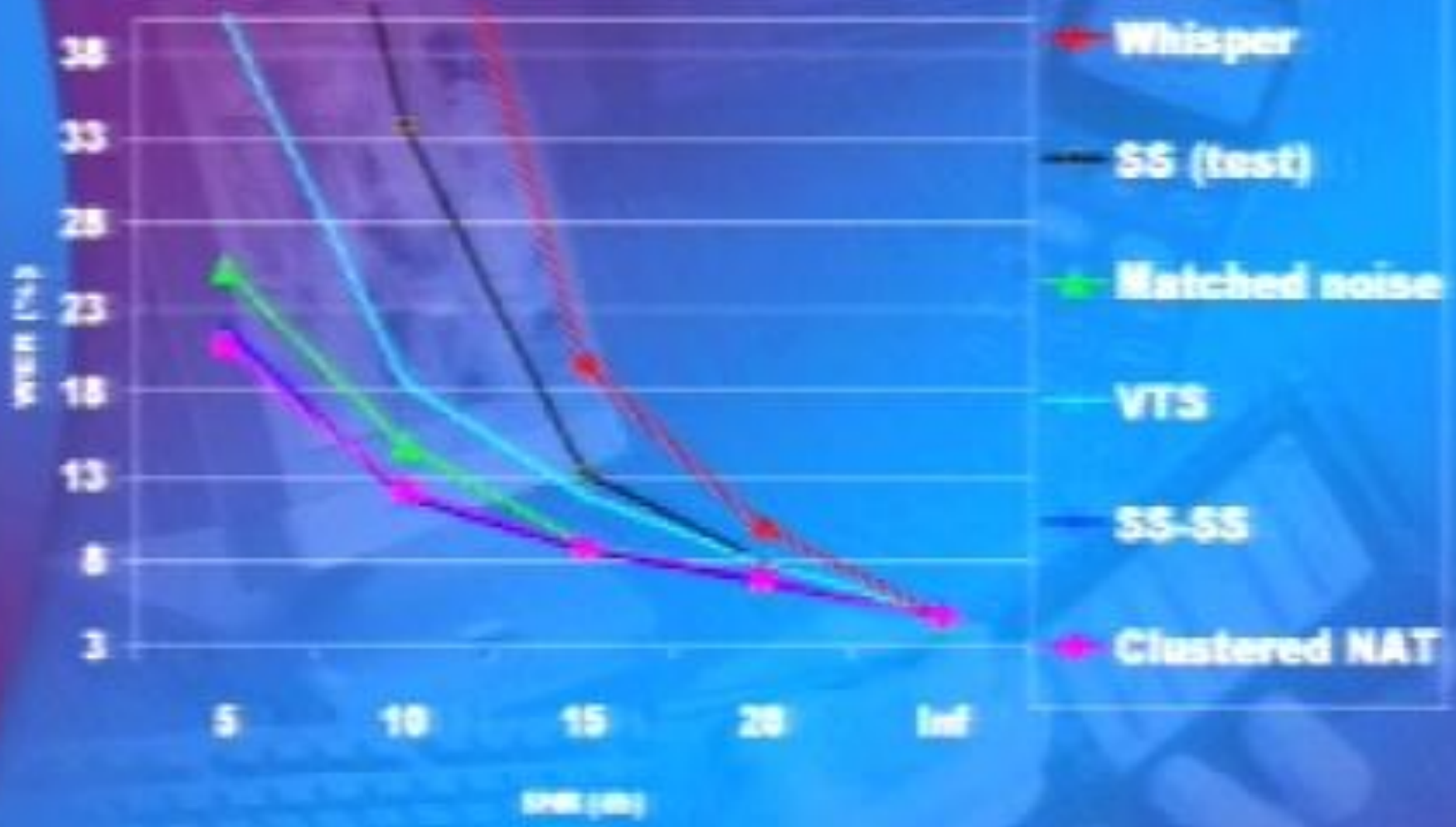
Preliminary Benchmark



Technologies: Noise Robustness

- Today's technology – 2 approaches exist to noise, noise reduction in the feature space and construction of models matching noisy test speech
- Concept - Noise Adaptive Training (NAT) combines above approaches
- Status – initial studies using white noise show great promise, studies using more realistic noise are underway. Plan to integrate into Mipad this summer

White Noise Results



MiPad Built-in Microphone

Technologies: Unified Engine

- Today's technology – recognizer and parser use separate language models
- Concept – the recognizer should use the parser language model (a well-defined grammar) to improve its spoken language understanding performance
- Status – offline studies underway, plan to integrate into Mipad this summer

Technologies: Dialog Management

- Today's technology – grammars must be explicitly defined, context plays little part in grammar
- Concept – Semantic classes provide a degree of abstraction for grammar authors, dialog memory provides context
- Status – integration into Mipad underway

Technologies: Dialog Management

- Today's technology – grammars must be explicitly defined, context plays little part in grammar
- Concept – Semantic classes provide a degree of abstraction for grammar authors, dialog memory provides context
- Status – integration into Mipad underway

Technologies: Small Devices

- Today's technology – command and control recognizers are compact, but dictation recognizer is too large to fit on mobile devices
- Status – Mipad release one will use a client-server architecture to overcome mobile device performance limitations. Research is underway to rectify this

Blind Source Separation

- In a sound-proof room:
  - Left mic ( ) right mic ( )
  - algorithm offers best published results
  - English ( ), Spanish ( )
- In a normal room, it is not working yet

Mipad Futures

- Mipad is just the first step, future directions include:
  - External products based on this technology
  - Greater use of dialog and context
  - Hands-free, eyes-free applications using speech output
  - Scaling the Mipad server up to a general speech server for other applications

Major Competitors

- DoCoMo's I-Mode: ahead of Europe and USA
- Motorola's Accompli: available in China only
- IBM is to focus on smaller devices

Major Competitors

- DoCoMo's I-Mode: ahead of Europe and USA
- Motorola's Accompli: available in China only
- IBM is to focus on smaller devices

Speech in your products

- Distributed speech allows products to take advantage of speech recognition without needing the necessary computing resources available locally
- Speech server efforts at MS will expand in the near future
- Local speech recognition is well supported in SAPI 5
- Small device recognition work ongoing

Shameless Plug

- Several reasons for this talk:
  - Technology transfer, stimulate ideas in the product groups
  - Get ideas and input on Mipad
  - Inform people about Mipad and the opportunities in STG
- We have job openings in STG!
  - Dev, test, PM, management all needed
  - Great technology and exciting environment
  - Contact myself or Xuedong Huang for more information


Where do you want to go
today?
Microsoft